Genomics Proteomics Bioinformatics 11 (2013) 172–181

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ORIGINAL RESEARCH

Identification of ta-siRNAs and *Cis*-nat-siRNAs in Cassava and Their Roles in Response to Cassava Bacterial Blight

Andrés Quintero, Alvaro L. Pérez-Quintero, Camilo López *

Departamento de Biología, Universidad Nacional de Colombia, Bogotá DC 111321, Colombia

Received 8 October 2012; revised 20 February 2013; accepted 14 March 2013
Available online 7 May 2013

KEYWORDS

ta-siRNA;
nat-siRNA;
Natural antisense transcript;
Cassava;
Xanthomonas

Abstract *Trans*-acting small interfering RNAs (ta-siRNAs) and natural *cis*-antisense siRNAs (*cis*-nat-siRNAs) are recently discovered small RNAs (sRNAs) involved in post-transcriptional gene silencing. ta-siRNAs are transcribed from genomic loci and require processing by microRNAs (miRNAs). *cis*-nat-siRNAs are derived from antisense RNAs produced by the simultaneous transcription of overlapping antisense genes. Their roles in many plant processes, including pathogen response, are mostly unknown. In this work, we employed a bioinformatic approach to identify ta-siRNAs and *cis*-nat-siRNAs in cassava from two sRNA libraries, one constructed from healthy cassava plants and one from plants inoculated with the bacterium *Xanthomonas axonopodis* pv. manihotis (*Xam*). A total of 54 possible ta-siRNA loci were identified in cassava, including a homolog of *TAS3*, the best studied plant ta-siRNA. Fifteen of these loci were induced, while 39 were repressed in response to *Xam* infection. In addition, 15 possible *cis*-natural antisense transcript (*cis*-NAT) loci producing siRNAs were identified from overlapping antisense regions in the genome, and were found to be differentially expressed upon *Xam* infection. Roles of sRNAs were predicted by sequence complementarity and our results showed that many sRNAs identified in this work might be directed against various transcription factors. This work represents a significant step toward understanding the roles of sRNAs in the immune response of cassava.

Introduction

Small RNAs (sRNAs) have emerged as important factors in the regulation of gene expression, which are involved in transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) through inactivation, degradation or translational repression of mRNAs [1,2]. One of the best known types of plant sRNAs is microRNAs (miRNAs). miRNAs originate from nuclear genomic loci and are transcribed by RNA polymerase II as primary miRNAs (pri-miRNAs), which are processed by Dicer like-1 (DCL1) to form mature miRNAs

* Corresponding author.
E-mail: celopezc@unal.edu.co (López C).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.gpb.2013.03.001

(21 nt) [1,2]. The mature miRNAs are then incorporated into an RNA-induced silencing complex (RISC), where they guide the degradation of complementary mRNAs [1,2].

Alternatively, miRNAs can target other non-coding RNAs (ncRNAs) to generate a type of plant-specific small interfering RNAs called *trans*-acting small interfering RNAs (ta-siRNAs) [3,4]. ta-siRNA loci (*TAS*) are transcribed from nuclear genes and then recognized and cleaved by Argonaute (AGO)-containing RNA-induced silencing complex (RISC) coupled with particular miRNAs [3,4]. The cleaved products are then used by an RNA-dependent RNA polymerase (RDR6) to produce double-stranded RNAs (dsRNAs), which are sequentially cleaved by DCL proteins to produce phased, ~20–24 nt ta-siRNAs [3,4]. These ta-siRNAs function similarly to miRNAs by targeting complementary genes and initiating their degradation or translation arrest [3,4]. Each *TAS* is cleaved by one or two miRNAs, producing several ta-siRNAs in phase and resulting in an amplified signal that silences several target molecules [3,4]. To date, 8 *TAS* loci have been identified and validated in *Arabidopsis*. These loci are grouped into 4 families, *TAS1–4*, each with a distinct biogenesis and function. The *TAS1* family is comprised of 3 highly similar genes that target 5 transcripts with unknown function [3,4]. The *TAS2* family contains ta-siRNAs that target mRNAs coding for pentatricopeptide repeat (PPR) proteins [4]. The cleavage of both *TAS1* and *TAS2* transcripts is directed by miR-173 in conjunction with AGO1 [3,5]. *TAS3* is highly conserved in plants, including monocots and dicots, and cleaved through the combined action of miR-390 and AGO7 [6–8], producing ta-siRNAs that target auxin-responsive factors (ARF) [6–8]. *TAS3*, its targets (ARF2–4) and miR-390 are involved in a self-regulatory loop, controlling lateral root growth [9,10]. In this loop, the ta-siRNAs produced from *TAS3* inhibit the expression of ARF2/3/4, ARF4 prevents miR-390 accumulation [9,10], while ARF3 up-regulates miR-390 accumulation in response to auxin [9,10]. The cleavage of *TAS*4 is triggered by miR-828, producing a series of ta-siRNAs that target several MYB transcription factors including PAP1, PAP2 and MYB113 [11,12].

In addition to miRNAs and ta-siRNAs, there are other gene-silencing mechanisms involving sRNAs, such as nat-siRNAs from natural antisense transcripts (NATs), which are pairs of endogenous coding or non-coding RNAs with perfect complementarity to each other [13]. There are two classes of nat-siRNAs: *cis* and *trans*. The former are derived from NATs located in the opposite strand of the same genomic locus. They originate when both transcripts in the NAT-pair are simultaneously transcribed, and their complementary regions match up forming a dsRNA, which is employed as a template to produce siRNAs [14]. The siRNA processing is dependent on DCL1 and/or DCL3, the siRNAs are predominantly produced from only one of the transcripts and they direct the cleavage of the transcript produced in the opposite strand. In this manner the transcripts involved in the *cis*-NAT pair are reciprocally regulated [15]. On the other hand, *trans*-nat-siRNAs arise from two overlapping transcripts that are located in different loci, and are processed in a similar manner as *cis*-nat-siRNAs [13].

The functions of ta-siRNA and *cis*-nat-siRNAs are largely unknown. However, some aspects of their overall importance in various plant processes have been established. A crucial role for ta-siRNAs has been reported for example in pollen maturation [16], in response to low temperature [17] and hypoxia [18]. Likewise, *cis*-nat-siRNAs play an important role in regulating the expression of the genes from which they arise; that is, if one of the genes is induced, the other is repressed [14]. This phenomenon has been observed in a wide variety of gene pairs with various functions [14]. For example, it has been shown in *Arabidopsis* that a sperm-specific *cis*-NAT pair, formed by KPL and its inversely transcribed gene ARIADNE14, is regulated by nat-siRNAs, which is required for the fertilization process. In the absence of KPL, no nat-siRNAs are produced, resulting in the induced regulation of ARIADNE14 RNA, increasing its levels and in consequence inhibiting fertilization [19].

While, some plant miRNAs have been found to have a direct role in defense against bacteria, fewer studies have focused on the role of ta-siRNAs or *cis*-nat-siRNAs. It is known, for example, that upon microbe-associated molecular pattern (MAMP)-triggered immunity (MTI), miR-393 is induced after recognition of the MAMP flagelin by the FLS2 receptor, which consequently targets ARFs and disrupts auxin signaling [20]. Additionally, other miRNA families also contribute to antibacterial defense by modulating hormonal responses [21,22].

There are, however, some indications that ta-siRNAs and nat-siRNAs may also be important for antibacterial defense. It was shown that miRNAs could target the tobacco *N* resistance gene and produce ta-siRNAs [23]. In addition, several ta-siRNAs target and regulate the expression of multiple NB-LRR-coding resistance genes in legumes during their interaction with *Rhizobium* [24]. On the other hand, siRNAs derived from a *cis*-NAT locus are induced in *Arabidopsis* in response to the effector AvrRpt2 from *Pseudomonas syringae* pv. tomato DC3000 [25].

Cassava (*Manihot esculenta*) is a staple crop that represents the main source of carbohydrates for more than 600 million people worldwide. Cassava starch also has a high potential for bioethanol production. Cassava is grown mainly by farmers in marginal areas of South America, Africa and Asia, where its production is severely affected by the bacterium *Xanthomonas axonopodis* pv. manihotis (*Xam*) [26]. Previously, we reported the identification of cassava miRNAs that were induced or repressed after inoculation with *Xam* through the sequencing of sRNA libraries [27]. Here, we expand our study on bacterial response of cassava by examining the function of small ncRNAs through identification of differentially expressed ta-siRNA and *cis*-nat-siRNAs.

Results

Identification of conserved ta-siRNAs in cassava

With the aims of establishing a set of parameters suitable for ta-siRNA prediction, the expression profiles of known *Arabidopsis TAS* loci and their conserved counterparts in cassava were analyzed.

We used the reported *Arabidopsis TAS* sequences as a query to screen the cassava genome for loci that have a similar sequence and are also predicted to have miRNA cleavage sites. Only one locus showed high similarity to an *Arabidopsis* locus: *TAS3*. The overall nucleotide identity was 54% over the length of the *Arabidopsis* locus and the location of this conserved *TAS* in the latest version of the cassava genome was scaffold00708:185488,185707. From now on, this sequence is

referred to as *CassTAS3*. In this locus, two recognition sites for cassava miR-390 homologs were identified: CassmiR-390a and CassmiR-390b (miRNA annotation as presented in [27]). This finding was in agreement with the predicted cleavage sites for *TAS3* family members in other plant species [3]. Aligned reads to the cassava genome in the *CassTAS3* locus showed the presence of abundant phased sRNAs approximately 21 nt long in the region between the two predicted cleavage sites for miR-390 (**Figure 1**). Similar expression profiles were obtained for the *Arabidopsis TAS* loci (including *TAS3*), when sRNA reads from available libraries (16 in total) were aligned against the *Arabidopsis* genome (Figure 1). These data suggest that a *TAS3* homolog exists in cassava and that it is transcribed and processed in a similar manner as in *Arabidopsis*.

Identification of novel ta-siRNAs in cassava

Based on these findings, a methodology was designed to identify novel *TAS* loci producing phased sRNAs, relying on expression profiles and the presence of miRNA cleavage sites. For this, reads from the cassava sRNA library were mapped to the cassava genome. The mapping results were screened using a 500-nt sliding window, to identify genomic regions where $\geqslant 3$ different mapped positions were found for sRNA reads, and where each mapped position was separated from the next one by

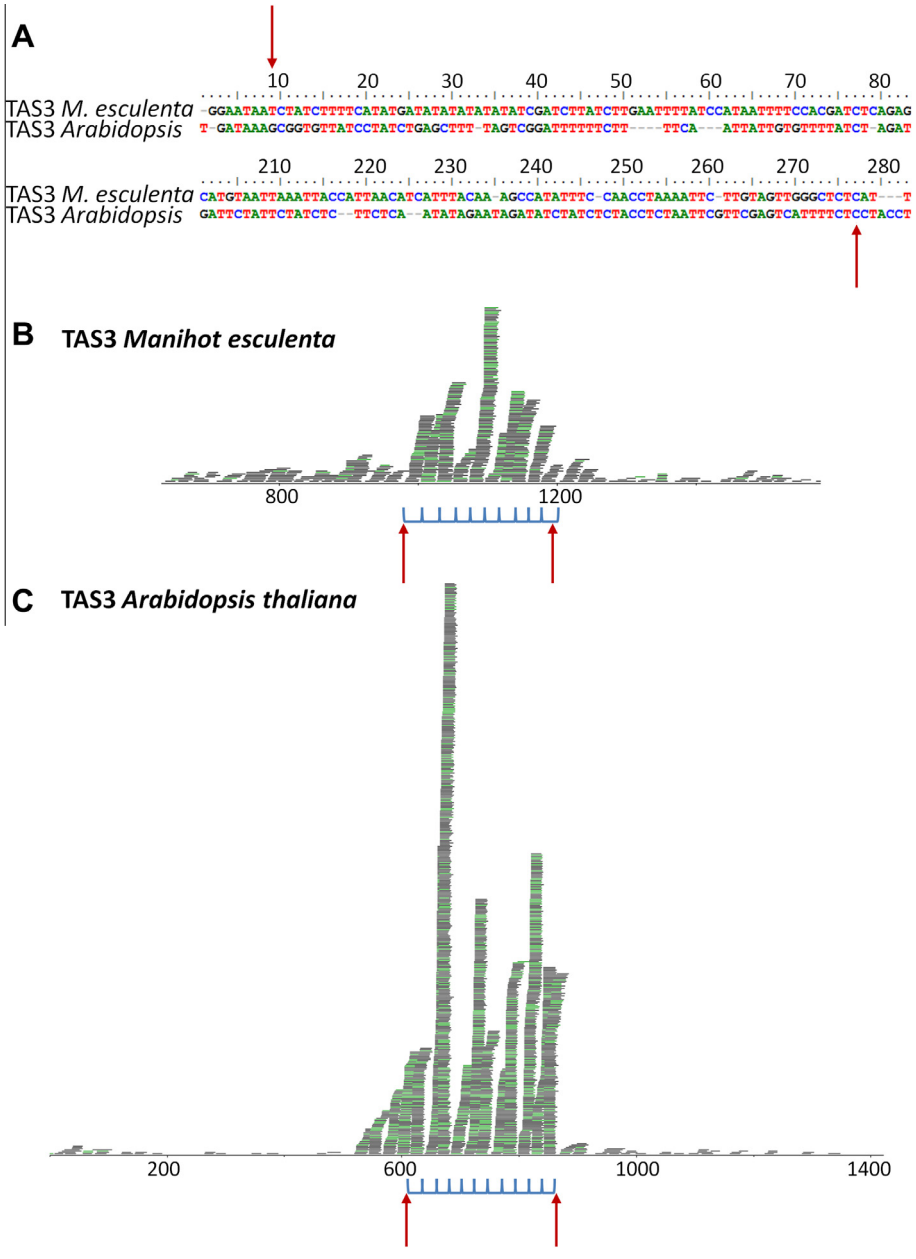

Figure 1 ***TAS3* is conserved in cassava**

A. Alignment between *A. thaliana TAS3* and *M. esculenta TAS3* (54% identity). **B.** sRNA reads mapped to *M. esculenta TAS3*. **C.** sRNA reads mapped to *A. thaliana TAS3*. Sense and antisense reads are shown in gray and antisense reads in green. Arrows indicate cleavage site by miR-390.

< 24 nt (suggesting a possible phased locus). Regions with these characteristics were retained as candidate phased siRNA loci, delimited from 50 nt downstream of the first mapped position to 50 nt upstream of the last mapped position. Regions mapped to known coding regions were eliminated. Then cleavage sites for cassava miRNAs were predicted in these candidate loci. In total, 239 candidate loci contained predicted binding sites for cassava miRNAs and produced phased sRNAs, these were kept as putative ta-siRNAs. Notably, 38 (16%) of these loci had putative cleavage sites for miR-156 and 12 (5%) loci have cleavage sites for miR-393. The process of ta-siRNA prediction is summarized in Figure S1.

To validate that these putative ta-siRNAs were independent transcriptional units producing sRNAs, mapping results were analyzed to identify sRNA enrichment in the candidate *TAS* loci. For this, reads from the sRNA libraries were mapped onto the candidate loci and their neighboring region (2 locus length upstream and downstream). The ratio of the expression between the candidate ta-siRNA and its neighboring region (candidate ta-siRNA/neighboring region) was calculated. The threshold for sRNA enrichment using this expression ratio was set as 4, according to the mapping results from Cass*TAS*3 and the *Arabidopsis TAS* loci. Out of 239 loci, 53 showed the expression ratio >4, indicating an enrichment of siRNA in these loci (Figure S2). These were considered putative cassava ta-siRNAs and are named Cass_tas01 to Cass-tas053 (Table S1).

Positive and negative controls for ta-siRNA identification

As a positive control, the same methodology was implemented in *Arabidopsis*, using sequences from the 16 sRNA libraries available. These analyses allowed the identification of 1020 candidate phased loci. Among them, 35 candidates had predicted target sites for *Arabidopsis* miRNAs including 8 ta-siRNAs reported for this species and other 27 loci (Table S2). These additional loci could be false positives but could also be previously unidentified *bona-fide* ta-siRNAs, and will thus be analyzed in subsequent works.

Additionally, two strategies were devised as negative controls to assess the impact of two possible sources of error using the data from *Arabidopsis*. First, it is possible that the phased loci found could be the result of random or unspecific mapping of sRNA reads to some genomic regions. To address this possibility, the sRNA sequences from the set of 16 libraries from *Arabidopsis* were shuffled and re-mapped to the *Arabidopsis* genome, the mapping results were analyzed to look for candidate phased loci using the same criteria mentioned above. Our results showed that none of the regions mapped with shuffled reads satisfied these criteria, and more importantly, the procedure was repeated 10 times (10 sets of shuffled reads) and we got the same result. Therefore, obtaining a phased locus from random mapping seems unlikely. Second, it is possible that the prediction of cleavage sites for miRNAs in the candidate phased loci may be due to fortuitous sequence complementarity. To address this possibility, 100 sets of shuffled miRNA sequences were generated from the set of *Arabidopsis* miRNAs and target predictions were performed against the set of candidate phased loci. As a result, some associations between random miRNAs and candidate *TAS* loci were predicted but were always fewer than those predicted with the original miRNA set. An average of 8 out of 1020 *TAS* candidates (min 3, max 11) were predicted as targets for random miRNAs, compared to 35 candidates with the original *Arabidopsis* miRNAs. Similar results were obtained when random miRNA sequences were generated from cassava miRNAs and screened against candidate cassava *TAS* loci. This indicates that chance miRNA-*TAS* loci associations may occur albeit with low frequency, thus experimental validation for miRNA cleavage is eventually needed for the *TAS* candidate loci presented in this work.

ta-siRNA expression in response to bacterial infection in cassava

To determine whether the candidate cassava ta-siRNAs are involved in responses to the bacteria *Xam*, the expression of cassava ta-siRNAs in the inoculated library was calculated (as normalized number of reads) and compared to that of the non-inoculated library. Expression of 13 ta-siRNAs was found to be repressed (log2fold change (FC) < −2, inoculated/non-inoculated) in response to the bacteria, while expression of 3 ta-siRNAs was induced (log2FC > 2) (Figure S3). The loci with the most repressed expression were Cass-tas31, Cass-tas28 and Cass-tas15 (**Figure 2**) and the loci with the most induced expression were Cass-tas42 and Cass-tas02 (**Figure 3**).

To identify the possible ta-siRNAs targets, sRNA reads derived from the identified *TAS* loci were screened against cassava genes using psRNATarget [28]. The targets of most of the ta-siRNAs were genes encoding proteins with an unknown function (32%) (Figure S4), which is consistent with previous studies [29]. A considerable number of ta-siRNA targets were proteins with LRR domains (17%), which are common in defense-related proteins [30], and PPR proteins (10%), which are known to be targets of other ta-siRNAs [31]. The conserved locus *CassTAS3* was predicted to target ARFs, as has been reported and validated previously [32].

Identification of *cis*-nat-siRNAs in cassava

Cassava genes were reciprocally aligned to find complementary (sense-antisense) regions. A total of 3000 gene pairs were identified with these characteristics (NATs). Of these, 128 genes were located in overlapping genomic regions and thus identified as *cis*-NAT loci. The expression of the genes involved in the NAT pair was verified by aligning them to a set of cassava ESTs.

To determine whether these loci were enriched in siRNAs, the profile of mapped reads against these genomic regions was analyzed. If *cis*-nat-siRNAs were produced from these loci, it is expected to find more mapped sRNA reads in the overlapping region of the NAT pair than in the coding regions of either gene, and ideally the mapped reads should come predominantly from one of the strands. Accordingly, the expression of sRNAs in the overlapping region of the complementary genes was compared to that of the non-overlapping region of the genes. Based on results obtained in *Arabidopsis* and *rice* [15], a threshold of 4 was established for the expression ratio (overlapping/non-overlapping) as a conservative parameter. 15 *cis*-NAT pairs showed sRNA enrichment in the overlapping region, suggesting the production of *cis*-nat-siRNAs (**Figure 4**). These loci were named as Cass_*cis*-nat01 to Cass_*cis*-nat15, respectively (Table S3). Among them, 13 of these loci showed strand bias (log_2fold (sense/antisense) > |1|) in the production of sRNAs in at least one of the libraries. The annotated possible functions of the gene pairs involved in the *cis*-NAT loci in most cases cor-

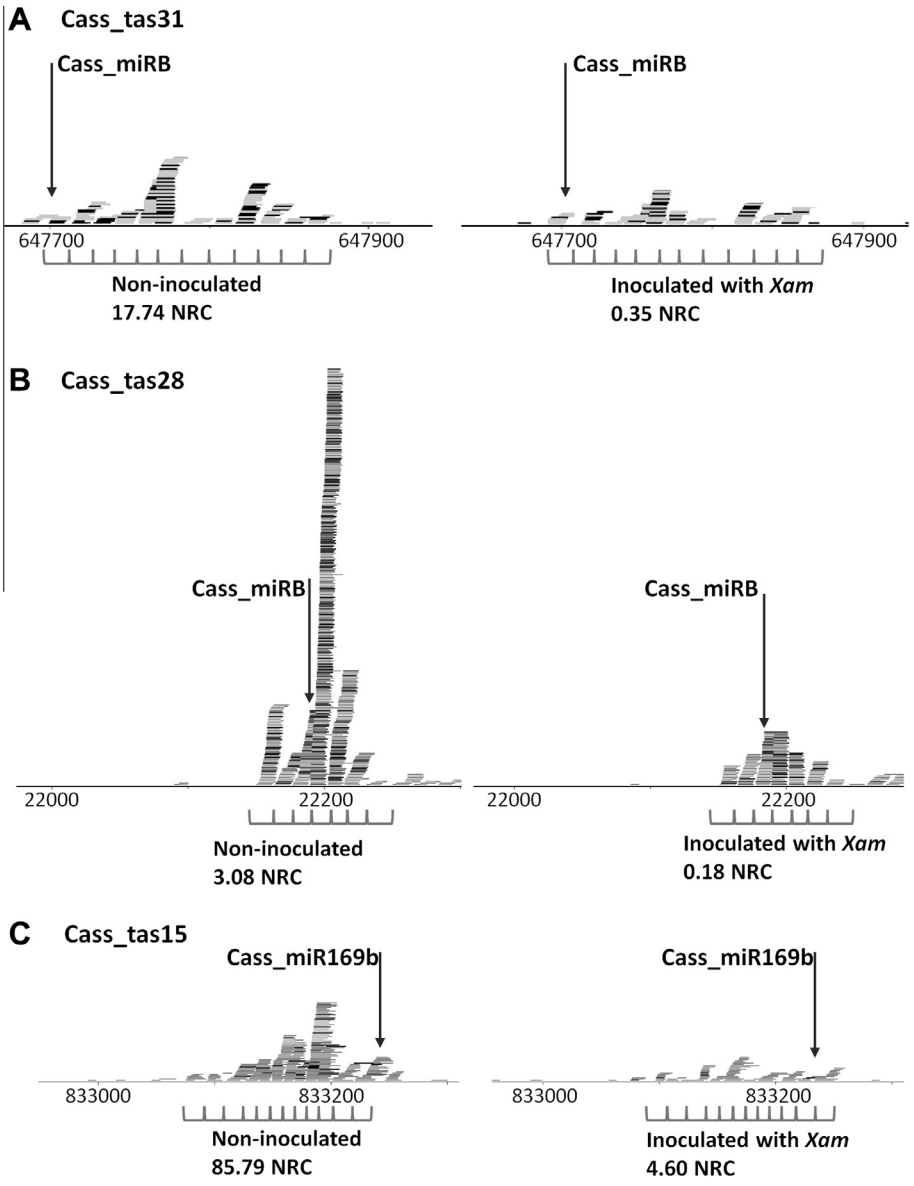

Figure 2 Bacteria-repressed ta-siRNA loci in cassava

sRNA mapping to new cassava TAs loci are shown for both *Xam*-inoculated and non-inoculated libraries, for significantly-repressed loci Cass_tas31 (**A**), Cass_tas28 (**B**) and Cass_tas14 (**C**), respectively. Sense reads are shown in light gray and antisense reads are shown in black. Arrows indicate cleavage site by the corresponding miRNA. Lines indicate clusters of phased-siRNAs. NRC stands for normalized read counts here.

responded to unrelated physiological or molecular processes. For example, Cass_*cis*-nat04 comprises the overlapping region between an Early Responsive to Dehydration 15 (ERD15) and a metal ion binding protein, with yet-unknown roles in cassava, but regulation of these genes by production of *cis*-nat-siRNAs may nonetheless be relevant in various physiological processes.

As with the *TAS* data, a strategy was devised to assess the possibility that sequences could map to these regions at random, the sequences of the cassava sRNA libraries were shuffled and mapped onto the genome. We found that no random sequences mapped to the overlapping region in *cis*-NAT pairs, the procedure was repeated 10 times, which produced the same results.

The gene expression ratio between the sense and antisense genes under both conditions (inoculated and non-inoculated) was also compared (Figure S5). Expression patterns for 6 *cis*-NAT candidates did not show a significant change. On the other hand, expression of 9 NAT candidates, including Cass_*cis*-nat04, showed sharp changes in response to *Xam*. In response to the bacteria, there is a larger production of sRNAs coming from the strand corresponding to the ERD15, which presumably silences the metal ion binding gene in the antisense strand. The relevance of this type of silencing to the plant-bacteria interaction remains to be studied.

Discussion

In this work, we have identified 54 cassava ta-siRNAs and found that expression of 3 ta-siRNAs was induced in response

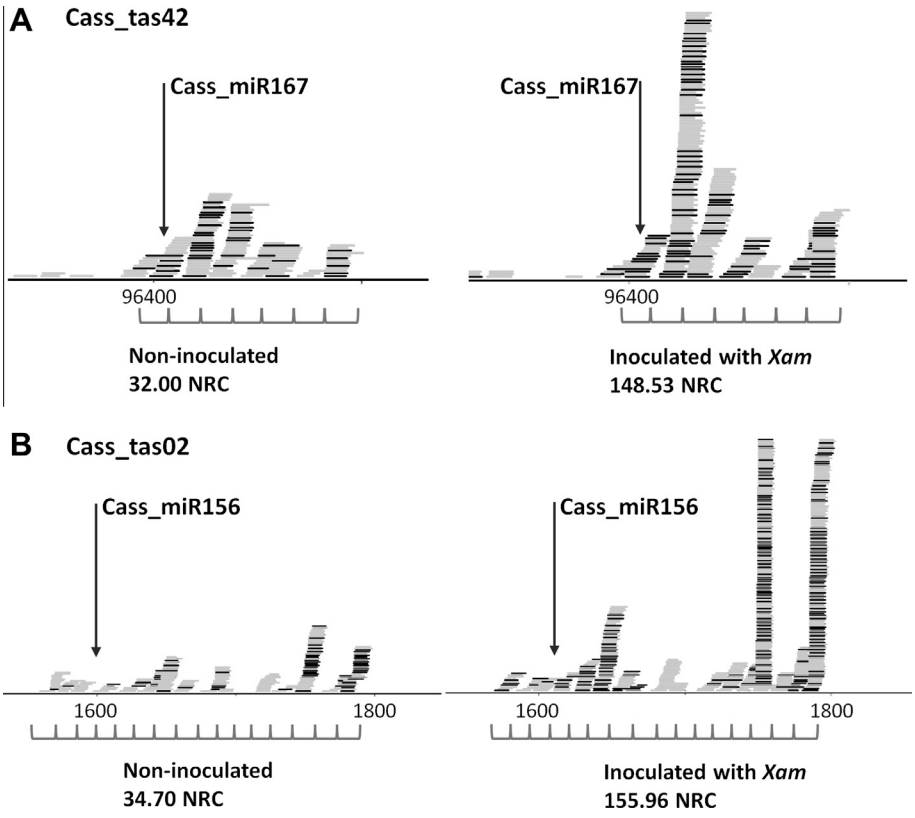

Figure 3 Bacteria-induced ta-siRNA loci in cassava
sRNA mapping to new cassava TAs loci are shown for both *Xam*-inoculated and non-inoculated libraries, for significantly-induced loci Cass_tas42 (**A**) and Cass_tas02 (**B**). Sense reads are shown in light gray and antisense reads are shown in black. Arrows indicate cleavage site by the corresponding miRNA. Lines indicate clusters of phased-siRNAs. NRC stands for normalized read counts here.

to bacterial infection, while expression of 13 ta-siRNAs was repressed. The targets of these ta-siRNAs comprised a diverse group of genes including ARFs.

To date, repertories of predicted ta-siRNAs or ta-siRNAs derived from direct sequencing have been reported for a limited number of plants, including *Arabidopsis* [6], rice [29], cotton [33], melon [34], legumes [24] and mosses [35]. The number of candidate ta-siRNAs identified in this study is significantly higher than the number of relatively well-known ta-siRNAs reported in *Arabidopsis*, which contains 8 ta-siRNA loci grouped in 4 classes (*TAS1* to *TAS4*). However, *TAS* loci may be underrepresented in *Arabidopsis*, given that hundreds of phased sRNAs have been reported in other species [29]. In this work, while validating the search strategy for ta-siRNAs, we found 25 possible new *TAS* loci from *Arabidopsis* which will be the subject of further studies. These results suggest that plants contain a larger repertoire of ta-siRNAs than initially anticipated and that this pathway is an important regulatory circuit for amplifying gene silencing.

Some ta-siRNAs are known to be involved in abiotic stress response [17,18] or in developmental processes [16]. In this study, the importance of certain ta-siRNAs in antibacterial defense is highlighted through the analysis of their expression in response to *Xam*. In our previous study, we reported that expression of several ARF-targeting miRNAs is induced in cassava in response to *Xam* [27]. In this work, *CassTAS*3, which is processed by the *Xam*-induced miR390, was also found targeting ARFs, although the locus expression was not significantly increased. It has been reported that auxin signaling is repressed in plants in response to pathogens to restrict their growth [20]. Disruption of auxin signaling is an important mechanism for broad-spectrum resistance, given that pathogens secrete auxins during infection to render a plant vulnerable by loosening its cell wall [36].

Recently, it was shown that miRNAs are directed against transcripts of the NB-LRR immune receptor family and trigger the production of phased secondary siRNAs in tobacco and other Solanaceae species [23]. In particular, resistance gene *N*, which is responsible for tobacco's resistance to TMV, is targeted by phased siRNAs and co-expression of gene *N* and phased siRNAs compromises resistance [23]. Also in the case of *Rhizobium*, several miRNAs and ta-siRNAs are induced to target NB-LRR genes, blocking the immune response and probably facilitating symbiotic interactions [24]. In this study, the targets of cassava ta-siRNAs in response to *Xam* are particularly enriched in genes encoding LRR-containing proteins. This finding suggests a possible role for ta-siRNAs in triggering the immune response to bacteria in cassava.

In this work, 15 *cis*-NAT loci producing siRNAs were also identified, and at least 9 of these showed changes in siRNA production in response to bacteria, and even in some cases

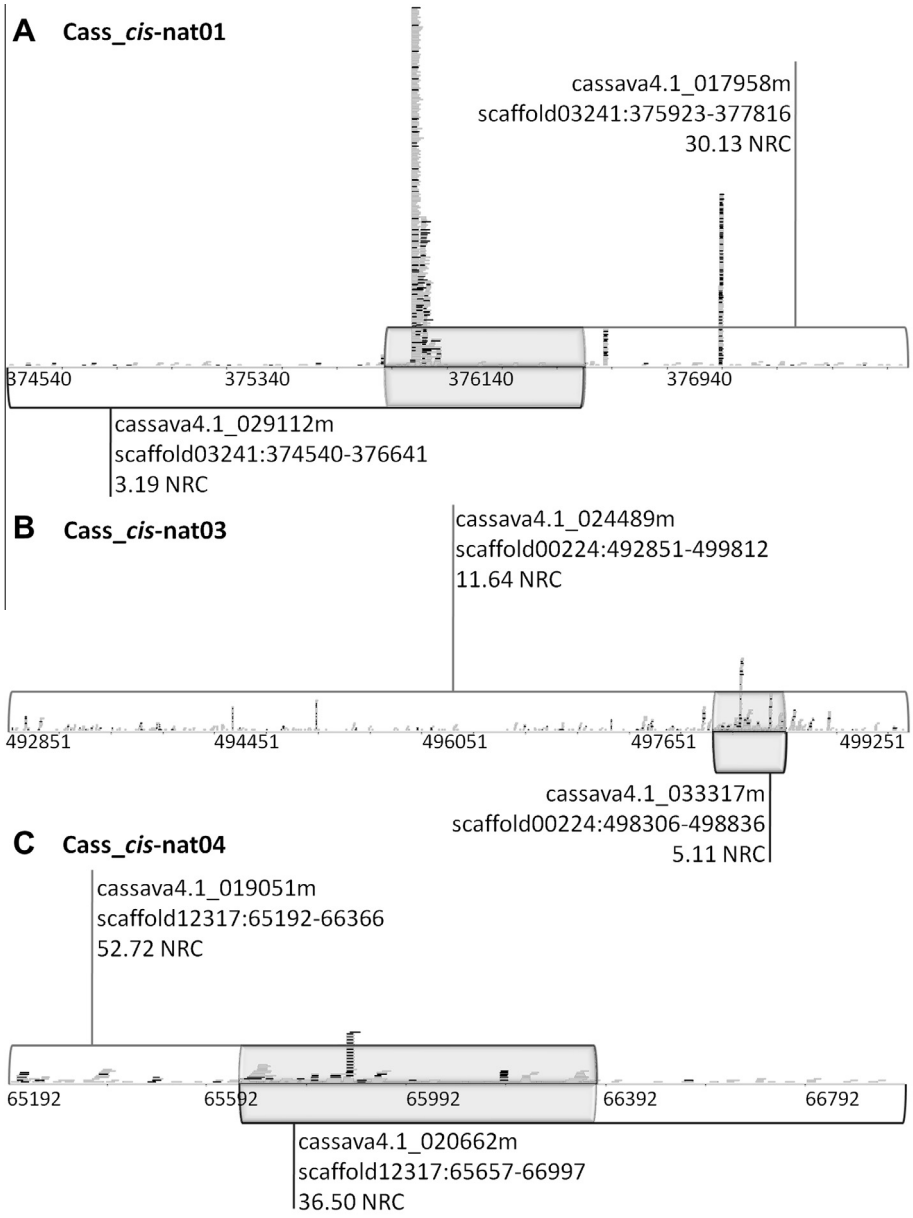

Figure 4 Representative cassava *cis*-NAT loci

Figure shows sRNA reads mapped to three representative *cis*-NAT loci. **A.** Cass-*cis*-nat01 comprises genes cassava4.1_017958m and cassava4.1_029112m. **B.** Cass_*cis*-nat03 comprises genes cassava4.1_024489m.g and cassava4.1_033317m.g. **C.** Cass_*cis*-nat04 comprises genes cassava4.1_019051m.g and cassava4.1_020662m.g. Small box indicates overlapping region between the two genes. NRC stands for normalized read counts here. Large boxes indicate the regions comprising the genes in the NAT loci, on the sense (upper box) and antisense (lower box) strand. Sense reads are shown in light gray and antisense reads are shown in black.

the production of siRNAs shifted from one strand to the other upon infection. The expression profiles of sRNA reads in these loci were of different nature, but the sRNAs were predominantly restricted to discrete sites in the overlapping region in the NAT pair. This phenomenon has also been observed in *Arabidopsis* and rice, which is believed to be due primarily to processing by DCL1 [15]. The functions of the gene pairs involved in the *cis*-NAT loci were often unrelated to plant defense and the outcome of the siRNA-mediated regulation of these genes for defense is hard to assess.

The identification of ta-siRNAs and *cis*-nat-siRNAs and their potential targets during cassava-*Xam* interactions contributes to our understanding of sRNA function and, in particular, of plant immune responses and represents a great contribution to the genomics of cassava, a plant with a recently sequenced genome where only few functional features are known. However there is still a lot of work to be done to assess the effect of sRNA regulation in response to bacteria. These results still need to be compared with transcriptome data and other experimental methods for further validation.

Materials and methods

Datasets

For *Arabidopsis thaliana* analyses, the genome sequence and *TAS* loci sequence, including AT2G27400, AT1G50055, AT2G39675, AT2G39681, AT3G17185, AT5G49615 and AT3G25795, were obtained from TAIR [37]. Sequences of known miRNAs were obtained from miRBase [38]. Available sRNA sequence libraries were obtained from the Gene Expression Omnibus (GEO) database at NCBI. These include GSM424742, GSM491572, GSM491573, GSM491574, GSM491575, GSM491576, GSM491577, GSM491578, GSM491579, GSM424743, GSM512702, GSM512703, GSM491567, GSM491568, GSM491569, GSM491570 and GSM491571.

For cassava analyses, two sRNA libraries were used (GSE29379), which have been fully described previously [27]. Cassava miRNA sequences used were also described previously [27]. Mappings were made using the latest version of the *M. esculenta* genome (Mesculenta_147) obtained from Phytozome (http://www.phytozome.net/cassava) and 81,726 ESTs available from NCBI's UniGene database (http://www.ncbi.nlm.nih.gov/unigene). Gene annotations for cassava were also obtained from the latest version of the *M. esculenta* genome from Phytozome (http://www.phytozome.net/cassava).

Identification of ta-siRNAs

To identify conserved ta-siRNAs, a blastn [39] search (parameters e-value 0.05 max_target_seqs 5 -best_hit_score_edge 0.05 -best_hit_overhang 0.25) was conducted using the 8 reported and validated *A. thaliana* ta-siRNAs against the cassava genome. Candidate loci were screened for miRNA targets using psRNATarget with default parameters [28]. Reads from cassava libraries were then mapped against the cassava genome using Bowtie [40] (parameters -n 0 -m 10 -l 15) and the expression profile of the candidate ta-siRNA was analyzed. The criteria to consider a locus as a true conserved ta-siRNA were as follows: (i) it shares similarity to a known *Arabidopsis* ta-siRNA with a maximum E-value of 0.001 and no gaps, and (ii) it shares a cleavage site for an miRNA that is homologous to a known site in *Arabidopsis*. Similar criteria were also implemented for *de novo* identification of ta-siRNAs. It was hypothesized that a ta-siRNA locus must (i) be enriched in phased reads from sRNA libraries, (ii) contain a predicted binding site for a known cassava miRNA near the locus, and (iii) correspond to a non-coding region of the genome.

Based on these criteria, a protocol was developed to identify novel ta-siRNA loci. For this, Bowtie mappings to non-coding regions of the genome (same parameters as above) were organized according to the mapped position. Using a sliding window of 500 nt, we looked for regions where at least 3 mapped positions for sRNA reads were found and where each mapped position was 24 nt (± 2 nt) apart from each other. Non-coding regions with mapped phased-siRNAs were then used to search for binding sites of known miRNAs as described above. These analyses were made using in-house perl scripts, which is available upon request.

To determine whether a candidate ta-siRNA locus was enriched in sRNA reads, the Bowtie mappings were used to calculate the expression of sRNAs in the candidate locus and the neighboring region (2 times the length of the candidate locus upstream and downstream). The expression levels of the candidate locus and the neighboring regions were calculated as normalized reads (reads mapped/total reads length of region mapped). An expression ratio of candidate locus/neighboring region was calculated. Only candidate loci with a ratio >4 were considered positive and further studied.

To determine differential expression of ta-siRNAs and to visually verify that siRNA were produced in phase, reads from each library were mapped against the candidate loci as described above. SAM files obtained from mapping were converted and visualized using SAMtools [41] and BamView [42].

All non-redundant reads mapping to a candidate ta-siRNA loci were used to screen for targets against annotated genes in the cassava genome. Target prediction was performed using psRNATarget with maximum E-value set as 5.0. This program is suitable to identify complementarity for different types of sRNAs [28] and has been successfully used for ta-siRNA target prediction [43].

Two strategies were used as negative controls with the *Arabidopsis* data. (1) 10 sets of random sRNA sequences were obtained by shuffling the sequences of the sRNA libraries using shuffleseq [44] and were mapped against the *Arabidopsis* genome. The mapping results were analyzed to search for loci satisfying the criteria for phased sRNA enrichment mentioned above, the number of loci resulting for each set of random sequences was registered. (2) 100 sets of random miRNA sequences were generated by shuffling the set of *Arabidopsis* miRNAs or cassava miRNAs. These sequences were used to predict binding sites in candidate *TAS* loci using psRNATarget with default parameters [28]. The number of *TAS* loci with predicted binding site was registered for each set and compared with the predictions with the original sets.

Identification of *cis*-nat-siRNAs

Candidate *cis*-nat-siRNAs loci were selected using the following criteria: (i) there must be an overlapping antisense region between two genes in their coding regions and (ii) the complementary region between the two genes must be enriched in reads from sRNA libraries [15]. Based on these criteria, a protocol was developed to identify *cis*-NAT loci in cassava. All cassava genes were mapped against all cassava antisense genes using BLASTN (parameters -strand minus e-value 0.05 max_target_seqs 5 -best_hit_score_edge 0.05 -best_hit_overhang 0.25) to find gene pairs with complementary regions between them. This mapping allows the identification of possible *trans*-NAT- and *cis*-NAT loci. However, due to the difficulties of assessing locus enrichment for *trans* interactions, only *cis* interactions were analyzed in this work. Gene pairs located in overlapping genomic regions were retained for further analysis.

To determine sRNA enrichment in candidate *cis*-NAT loci, reads from the cassava libraries were mapped onto the overlapping genomic regions of complementary gene pairs (*i.e.*, the candidate *cis*-NAT loci) and compared to neighboring non-overlapping regions. This mapping was made using Bowtie as described above [40]. The sRNA density for the regions was established as number of read/length of region and

the ratio of the sRNA densities in the overlapping (OR) region and non overlapping region (non-OR) was calculated [15]. Only candidate loci with a OR/non-OR ratio >4 were considered positive and further studied. Alignments were visualized as above. To determine co-regulation between gene pairs, the expression levels of the sense and antisense genes were obtained. Ratios were calculated for the inoculated and non-inoculated libraries.

For a negative control, the enrichment analysis described above was repeated using 10 sets of randomized sRNA sequences obtained using the program from the EMBOSS package shuffleseq [44]. For each set, the number of reads mapped to the overlapping regions of the candidate *cis*-NAT loci was registered and compared with the mappings using non-randomized reads.

Authors' contributions

AQ collected the data, performed the analyses and drafted the manuscript. ALPQ participated in strategy design and script writing for the analysis, and helped draft the manuscript. CL supervised the project, analyzed the results, drafted and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to thank Anika Steurer for her critical reading of this manuscript. This project was partially funded by Dirección de Investigaciones Sede Bogotá (Grant No. 201010011438) and Colciencias (Contract No. 221-2008).

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2013.03.001.

References

[1] Axtell MJ, Westholm JO, Lai EC. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol 2011;12:221.
[2] Li L, Liu Y. Diverse small non-coding RNAs in RNA interference pathways. Methods Mol Biol 2011;764:169–82.
[3] Allen E, Xie Z, Gustafson AM, Carrington JC. MicroRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. Cell 2005;121:207–21.
[4] Zhang C, Ng DW, Lu J, Chen ZJ. Roles of target site location and sequence complementarity in *trans*-acting siRNA formation in *Arabidopsis*. Plant J 2012;69:217–26.
[5] Montgomery TA, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, Sullivan CM, et al. AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A 2008;105:20055–62.
[6] Axtell MJ, Jan C, Rajagopalan R, Bartel DP. A two-hit trigger for siRNA biogenesis in plants. Cell 2006;127:565–77.
[7] Felippes FF, Weigel D. Triggering the formation of tasiRNAs in *Arabidopsis thaliana*: the role of microRNA miR173. EMBO Rep 2009;10:264–70.
[8] Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, et al. Specificity of argonaute7-miR390 interaction and dual functionality in *TAS3 trans*-acting siRNA formation. Cell 2008;133:128–41.
[9] Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, et al. MiR390, *Arabidopsis TAS3* tasiRNAs, and their auxin response factor targets define an autoregulatory network quantitatively regulating lateral root growth. Plant Cell 2010;22:1104–17.
[10] Yoon EK, Yang JH, Lim J, Kim SH, Kim SK, Lee WS. Auxin regulation of the microRNA390-dependent transacting small interfering RNA pathway in *Arabidopsis* lateral root development. Nucleic Acids Res 2010;38:1382–91.
[11] Hsieh LC, Lin SI, Shih AC, Chen JW, Lin WY, Tseng CY, et al. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. Plant Physiol 2009;151:2120–32.
[12] Luo QJ, Mittal A, Jia F, Rock CD. An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in *Arabidopsis*. Plant Mol Biol 2012;80:117–29.
[13] Jin H, Vacic V, Girke T, Lonardi S, Zhu JK. Small RNAs and the regulation of *cis*-natural antisense transcripts in *Arabidopsis*. BMC Mol Biol 2008;9:6.
[14] Zhou X, Sunkar R, Jin H, Zhu JK, Zhang W. Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. Genome Res 2009;19:70–8.
[15] Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, Gao S, et al. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol 2012;13:R20.
[16] Grant-Downton R, Le Trionnaire G, Schmid R, Rodriguez-Enriquez J, Hafidh S, Mehdi S, et al. MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*. BMC Genomics 2009;10:643.
[17] Kume K, Tsutsumi K, Saitoh Y. TAS1 *trans*-acting siRNA targets are differentially regulated at low temperature, and TAS1 *trans*-acting siRNA mediates temperature-controlled At1g51670 expression. Biosci Biotechnol Biochem 2010;74:1435–40.
[18] Moldovan D, Spriggs A, Yang J, Pogson BJ, Dennis ES, Wilson IW. Hypoxia-responsive microRNAs and *trans*-acting small interfering RNAs in *Arabidopsis*. J Exp Bot 2010;61:165–77.
[19] Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S. Proper regulation of a sperm-specific *cis*-nat-siRNA is essential for double fertilization in *Arabidopsis*. Genes Dev 2010;24:1010–21.
[20] Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, et al. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 2006;312:436–9.
[21] Zhang W, Gao S, Zhou X, Chellappan P, Chen Z, Zhang X, et al. Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. Plant Mol Biol 2011;75:93–105.
[22] Li Y, Zhang Q, Zhang J, Wu L, Qi Y, Zhou JM. Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. Plant Physiol 2010;152:2222–31.
[23] Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, et al. MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci U S A 2012;109:1790–5.
[24] Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, et al. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, *trans*-acting siRNAs. Genes Dev 2011;25:2540–53.

[25] Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas Jr A, Zhu JK, et al. A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci U S A 2006;103:18002–7.
[26] López CE, Bernal AJ. Cassava bacterial blight: using genomics for the elucidation and management of an old problem. Trop Plant Biol 2012;5:117–26.
[27] Perez-Quintero AL, Quintero A, Urrego O, Vanegas P, Lopez C. Bioinformatic identification of cassava miRNAs differentially expressed in response to infection by *Xanthomonas axonopodis* pv. manihotis. BMC Plant Biol 2012;12:29.
[28] Dai X, Zhao PX. PsRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 2011;39:W155–159.
[29] Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, et al. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 2009;19:1429–40.
[30] Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: shaping the evolution of the plant immune response. Cell 2006;124:803–14.
[31] Chen HM, Li YH, Wu SH. Bioinformatic prediction and experimental validation of a microRNA-directed tandem *trans*-acting siRNA cascade in *Arabidopsis*. Proc Natl Acad Sci U S A 2007;104:3318–23.
[32] Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, et al. Regulation of auxin response factor3 by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. Curr Biol 2006;16:939–44.
[33] Yin Z, Li Y, Han X, Shen F. Genome-wide profiling of miRNAs and other small non-coding RNAs in the *Verticillium dahliae*-inoculated cotton roots. PLoS One 2012;7:e35765.
[34] Gonzalez-Ibeas D, Blanca J, Donaire L, Saladie M, Mascarell-Creus A, Cano-Delgado A, et al. Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. BMC Genomics 2011;12:393.
[35] Talmor-Neiman M, Stav R, Klipcan L, Buxdorf K, Baulcombe DC, Arazi T. Identification of *trans*-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J 2006;48:511–21.
[36] Fu J, Liu H, Li Y, Yu H, Li X, Xiao J, et al. Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. Plant Physiol 2011;155:589–602.
[37] Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 2012;40:D1202–1210.
[38] Griffiths-Jones S. miRBase: microRNA sequences and annotation. Curr Protoc Bioinformatics 2010; Chapter 12:Unit 12.9.1–10. http://dx.doi.org/10.1002/0471250953.bi1209s29.
[39] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
[40] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[41] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9.
[42] Carver T, Harris SR, Otto TD, Berriman M, Parkhill J, McQuillan JA. BamView: visualizing and interpretation of next-generation sequencing read alignments. Brief Bioinform 2013;14:203–12.
[43] Rajeswaran R, Aregger M, Zvereva AS, Borah BK, Gubaeva EG, Pooggin MM. Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. Nucleic Acids Res 2012;40:6241–54.
[44] Rice P, Longden I, Bleasby A. EMBOSS: the european molecular biology open software suite. Trends Genet 2000;16:276–7.